

# Markov chain methods for small-set expansion

Ryan O’Donnell\*

David Witmer†

March 2, 2013

## Abstract

Consider a finite irreducible Markov chain with invariant distribution  $\pi$ . We use the inner product induced by  $\pi$  and the associated heat operator to simplify and generalize some results related to graph partitioning and the small-set expansion problem. For example, Steurer showed a tight connection between the number of small eigenvalues of a graph’s Laplacian and the expansion of small sets in that graph. We give a simplified proof which generalizes to the nonregular, directed case. This result implies an approximation algorithm for an “analytic” version of the Small-Set Expansion Problem, which, in turn, immediately gives an approximation algorithm for Small-Set Expansion. We also give a simpler proof of a lower bound on the probability that a random walk stays within a set; this result was used in some recent works on finding small sparse cuts.

## 1 Overview

Graph partitioning using spectral methods has recently been the subject of intensive study. Many results in this area have been proven using discrete-time random walks. However, these techniques work best when applied to regular graphs with nonnegative eigenvalues. As a result, it has become standard to move to a lazy version of a graph by adding self-loops, i.e. using  $(I + K)/2$  instead of  $K$  as the adjacency matrix. Much work has also focused on regular graphs only or considered the normalized Laplacian  $D^{-1/2}LD^{-1/2}$ .

In this work we show that these problems can be avoided using Markov chain techniques, leading to simpler and more general proofs of results related to spectral graph partitioning. Rather than using discrete-time random walks, we consider continuous-time random walks and the associated heat operator. “Smoothing out” the random walk makes the eigenvalues nonnegative, avoiding the need to move to lazy graphs and allowing our techniques to be directly applied to the original instance. In addition, we use the inner product defined with respect the invariant distribution  $\pi$  of the Markov chain representing a random walk on the graph. We are then able to use our methods directly on nonregular graphs.

We will now give a brief description of some previous results in spectral graph partitioning. Let  $G = (V, E)$  be a graph on  $n$  vertices. Let  $K$  be its (normalized) adjacency matrix, let  $L$  be its (normalized) Laplacian matrix (namely  $I - K$ ), and let  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$  be the eigenvalues of  $L$ . The conductance  $\Phi[S]$  of a set  $S \subseteq V$  is defined to be  $\frac{E(S, \bar{S})}{\sum_{v \in S} \deg(v)}$ . The conductance profile of  $G$ , denoted  $\Phi_G$ , was defined by Lovász and Kannan [LK99] as

$$\Phi_G(r) = \min\{\Phi[S] : S \subseteq V, \mu[S] \leq r\}.$$

---

\*Department of Computer Science, Carnegie Mellon University. Supported by NSF grants CCF-0747250 and CCF-0915893, and by a Sloan fellowship.

†Department of Computer Science, Carnegie Mellon University.

Cheeger’s inequality for graphs [AM85, Alo86, SJ89] states that  $V$  can be partitioned into nonempty  $S_1, S_2$  such that  $\Phi[S_i] \leq O(\sqrt{\lambda_2})$ . Very recently, Louis, Raghavendra, Tetali, and Vempala [LRTV12] and Lee, Oveis Gharan, and Trevisan [LOT12] have given a “higher order Cheeger inequality” involving higher eigenvalues. Specifically, the two results show that for any  $k$ , one can partition  $V$  into  $\Omega(k)$  disjoint nonempty sets  $S_i$ , each of which has conductance  $\Phi[S_i] \leq O(\sqrt{\lambda_k \log k})$ . Since one of these parts has volume  $\mu[S_i] := |S_i|/|V| \leq O(1/k)$  we may conclude that

$$\Phi_G(\frac{\text{const}}{k}) \leq O(\sqrt{\lambda_k \log k}). \quad (1)$$

As noted in these works, for a fixed  $k$  the “extra factor” of  $\Theta(\sqrt{\log k})$  in (1) is necessary; indeed this is true [LOT12] for all  $k \leq \log_2 n$ . However, somewhat intriguingly, the extra factor becomes *unnecessary* once  $k$  is as large as  $n^{\Omega(1)}$  — at least, if one is willing to compromise somewhat on the volume parameter. Specifically, Arora, Barak, and Steurer [ABS10] showed for regular graphs that

$$\Phi_G(O(k^{-1/100})) \leq O(\sqrt{\lambda_k \log_k n}). \quad (2)$$

In his thesis, Steurer [Ste10] improved this bound to

$$\Phi_G(k^{-1+1/A}) \leq O(\sqrt{A \lambda_k \log_k n}) \quad \text{for any (sufficiently large) constant } A. \quad (3)$$

Using Markov chain methods, we give what we feel is a much simpler proof of this result, which also works for the nonregular (and also directed) case. Our result also implies an approximation algorithm for an “analytic” version of the Small-Set Expansion problem. This, in turn, immediately gives an approximation algorithm for Small-Set Expansion by a standard version of Cheeger’s Inequality,

In somewhat related recent work, Oveis Gharan and Trevisan [OT12] proved a weaker version of this bound with  $k^{-1/3}$  in place of  $k^{-1+1/A}$ . The main point of that work, along with the independent work of Kwok and Lau [KL12] give a polynomial-time algorithm for the Small-Set Expansion problem in an unweighted (nonregular) graph  $G = (V, E)$  with the following guarantee: if there exists  $S \subseteq V$  with  $\mu[S] \leq \delta$  and  $\Phi[S] \leq \epsilon$ , the algorithm finds  $T \subseteq V$  with  $\mu[T] \leq O(\delta) \cdot (\delta|E|)^\alpha$  and  $\mu[S] \leq O(\sqrt{\epsilon/\alpha})$  (for any small  $\alpha > 0$ ). To achieve this, both papers prove a theorem stating that for any  $S \subseteq V$  and integer  $t > 0$ , the probability that a  $t$ -step random walk starting from a random  $x \in S$  stays entirely within  $S$  is at least  $\left(1 - \frac{\Phi[S]}{2}\right)^t$ . We also give a simpler proof of this result for continuous-time random walks.

## 1.1 Our results

### 1.1.1 Bounding the spectral profile

In this work we provide a different, simple proof of Steurer’s improved result using continuous-time random walks instead of lazy discrete-time random walks:

**Theorem 1.1.** *In any strongly connected graph  $G$ ,  $\Phi_G(16k^{-1+1/A}) \leq 2\sqrt{A} \cdot \sqrt{\lambda_k \log_k n}$  for any real  $A \geq 3$ .*

For example,  $\Phi_G(k^{-.999}) \leq O(\sqrt{\lambda_k \log_k n})$  for  $k$  sufficiently large. See Section 2 for the appropriate definitions of  $\Phi_G$ ,  $L$ ,  $\lambda_i$ , etc. in the context of general graphs  $G$ .

In fact, our result is stronger than this in that we are able to directly bound the *spectral profile* of  $G$ . (The same is true of the result in Arora–Barak–Steurer [ABS10] and in Steurer’s

thesis [Ste10].) Recall that the spectral profile  $\Lambda_G$  of  $G$ , introduced by Goel, Montenegro, and Tetali [GMT06], is defined by

$$\Lambda_G(r) = \min \left\{ \frac{\langle f, Lf \rangle}{\|f\|_2^2} : \text{nonzero } f : V \rightarrow \mathbb{R}^{\geq 0} \text{ with } \pi(\text{supp}(f)) \leq r \right\}.$$

Goel, Montenegro, and Tetali showed that the “Cheeger rounding analysis” yields the following relationship with conductance profile:  $\Phi_G(r) \leq \sqrt{2\Lambda_G(r)}$  for all  $r$ .<sup>1</sup> As in [ABS10] we work with a slightly different definition of spectral profile, for technical convenience:

$$\Lambda'_G(r) = \min\{\Phi[f] : \mu[f] \leq r\}, \quad \text{where} \quad \Phi[f] = \frac{\langle f, Lf \rangle}{\|f\|_2^2}, \quad \mu[f] = \frac{\|f\|_1^2}{\|f\|_2^2}$$

are appropriate generalizations of boundary size and volume to functions  $f : V \rightarrow \mathbb{R}$ . (These definitions agree with our earlier ones when  $f$  is the 0-1 indicator of a set  $S \subseteq V$ .) As noted in [ABS10, Lemma A.2] we have  $\Lambda_G(4r) \leq 2\Lambda'_G(r)$  for all  $r$ . (A similar reverse connection also holds.) Thus:

**Theorem 1.2.** (Essentially from [GMT06].)  $\Phi_G(4r) \leq 2\sqrt{\Lambda'_G(r)}$  for all  $r$ .

We use this connection to obtain Theorem 1.1; our main theorem is in fact:

**Theorem 1.3.** In any strongly connected graph  $G$ ,  $\Lambda'_G(4k^{-1+1/A}) \leq A \cdot \lambda_k \log_k n$  for any real  $A \geq 3$ .

This route to bounding the conductance profile is somewhat in contrast to the works [LRTV12, LOT12], both of which combine their spectral analysis and “rounding algorithm”.

Indeed, in this work we consider the “analytic” version of the Raghavendra–Steurer [RS10] Small-Set Expansion problem: given a graph  $G = (V, E)$  with the promise that there is a function  $f : V \rightarrow \mathbb{R}$  which has  $\mu[f] \leq \delta$  and  $\Phi[f] \leq \epsilon$ , find a function  $g : V \rightarrow \mathbb{R}$  with  $\mu[g] \leq O(\delta)$  and  $\Phi[g]$  as small as possible. Following [ABS10], we provide an eigenspace enumeration lemma which, when combined with Theorem 1.3, yields the following:

**Theorem 1.4.** For any  $\alpha \leq \frac{1}{3}$  and  $C \geq 1$ , there exists an algorithm running in time  $\exp(O(n^\alpha) \cdot \frac{1}{\delta} \log(C/\delta))$  with the following guarantee: If there exists  $f : V \rightarrow \mathbb{R}$  with  $\mu[f] \leq \delta \leq 1/2$  and  $\Phi[f] \leq \epsilon \leq 1/4$ , the algorithm finds  $g : V \rightarrow \mathbb{R}$  with  $\mu[g] \leq \delta \cdot (1 + 1/C)$  and  $\Phi[g] \leq O(\frac{C^2}{\alpha\delta}) \cdot \epsilon$ .

As a byproduct, using Theorem 1.2 we can immediately deduce the following approximation algorithm for Small-Set Expansion:

**Corollary 1.5.** Fix any small constants  $\alpha, \delta > 0$ . Then there is an algorithm running in time  $\exp(O(n^\alpha))$  with the following guarantee: If there exists  $S \subseteq V$  with  $\mu[S] \leq \delta$  and  $\Phi[S] \leq \epsilon$ , the algorithm finds  $T \subseteq V$  with  $\mu[T] \leq 5\delta$  and  $\Phi[T] \leq O(\sqrt{\epsilon})$ .

More generally, one can obtain  $\Phi[T] \leq O(\epsilon^{\beta/2})$  in time  $\exp(O(n^{\alpha\epsilon^{1-\beta}}))$  for any  $0 < \beta \leq 1$ .

This result is incomparable with the Arora–Barak–Steurer Small-Set Expansion algorithm: their work had  $O(\epsilon^{\beta/3})$  in place of  $O(\epsilon^{\beta/2})$  and was analyzed only for regular graphs. On the other hand, our Corollary 1.5 holds only for  $\delta$  a constant, whereas their algorithm works for  $\delta$  as small as  $n^{-\epsilon^{1-\beta}}$  (which is the more interesting parameter range).

---

<sup>1</sup>Actually, [GMT06] defined  $\Lambda_G(r)$  as the minimization of  $\frac{\langle f, Lf \rangle}{\|f\|_2^2 - \|f\|_1^2}$ . But their proof of this relationship still goes through.

### 1.1.2 Continuous-time random walks

In [OT12], Oveis Gharan and Trevisan prove a lower bound on the probability that a random walk stays within a set. (Kwok and Lau [KL12] prove a similar but somewhat weaker bound.) Specifically, they show:

**Theorem 1.6.** *Let  $G = (V, E)$  be an undirected graph with invariant distribution  $\pi$ . Let  $\emptyset \neq S \subseteq V$  and let  $t > 0$  be an integer. Choose  $\mathbf{x} \sim \pi$  conditioned on  $\mathbf{x} \in S$ , and then perform a  $t$ -step discrete-time random walk from  $\mathbf{x}$ . Then the probability that the walk stays entirely within  $S$  is at least  $\left(1 - \frac{\Phi[S]}{2}\right)^t$ .*

We provide a simple proof of a similar theorem using Markov chain methods.

**Theorem 1.7.** *In the setting of Theorem 1.6, if we instead perform a time- $t$  continuous-time random walk, the probability that the walk stays entirely within  $S$  is at least  $\exp(-t\Phi[S])$ .*

## 2 Preliminaries

Instead of directed graphs, we will use the language of Markov chains; for background, see e.g. [DSC96, MT06].

Throughout this work,  $G$  will denote an irreducible Markov chain on state space  $V$  of cardinality  $n$ , with no isolated states. We will be considering elements  $f$  in the vector space of functions  $V \rightarrow \mathbb{R}$ . We write  $K$  for the adjacency matrix operator:  $Kf(x) = \mathbf{E}_{\mathbf{y} \sim x}[f(\mathbf{y})]$ , where  $\mathbf{y} \sim x$  denotes that  $\mathbf{y}$  is obtained by taking one step from  $x$  in the chain.  $K$  has a unique invariant probability distribution  $\pi$  on  $V$  which is nowhere 0. It gives rise to an inner product on functions,  $\langle f, g \rangle = \mathbf{E}_{\mathbf{x} \sim \pi}[f(\mathbf{x})g(\mathbf{x})]$ . We write  $L = id - K$  for the Laplacian operator and  $H_t = \exp(-tL)$  for the heat kernel (continuous-time transition) operator.

**Definition 2.1.** Given nonzero  $f : V \rightarrow \mathbb{R}$  we define its *analytic boundary size/conductance* to be

$$\Phi[f] = \frac{\langle f, Lf \rangle}{\langle f, f \rangle} = 1 - \frac{\langle f, Kf \rangle}{\langle f, f \rangle}.$$

Note that if  $f$  is the 0-1 indicator of a set  $S \subseteq V$  then  $\Phi[f] = \Pr_{\mathbf{x} \sim \pi, \mathbf{y} \sim \mathbf{x}}[\mathbf{y} \notin S \mid \mathbf{x} \in S]$ . We will also write  $\Phi[S]$  in this case.

**Definition 2.2.** Given a nonzero  $f : V \rightarrow \mathbb{R}$  we define its *analytic sparsity* to be

$$\mu[f] = \frac{\|f\|_1^2}{\|f\|_2^2}.$$

Note that if  $f$  is the 0-1 indicator of a set  $S \subseteq V$  then  $\mu[f] = \pi(S)$ .

These definitions motivate consideration of an “analytic” version of the Small-Set Expansion Problem: Assuming there is an analytically sparse  $f$  with small analytic boundary, find such an  $f$ . More precisely:

**Analytic Small-Set Expansion Problem:** Given as input  $G$  with the promise that there exists  $f : V \rightarrow \mathbb{R}$  with  $\mu[f] \leq \delta \leq 1/2$  and  $\Phi[f] \leq \epsilon$ , find  $f' : V \rightarrow \mathbb{R}$  with  $\mu[f'] \leq \delta'$  and  $\Phi[f'] \leq \epsilon'$ . In this bicriteria problem, we typically insist that  $\delta' = O(\delta)$  and then try to minimize  $\epsilon'$ .

Note that the standard Small-Set Expansion problem is the above problem with the additional restriction that  $f$  and  $f'$  should be 0-1-valued functions.

For the remainder of this work we will assume that  $G$  is reversible. However, this is without loss of generality since, given a non-reversible Markov chain  $G'$  with adjacency matrix operator  $K'$ , we can replace it with the reversible Markov chain  $G$  having adjacency matrix operator  $K = \frac{K' + K'^*}{2}$ . The chain  $G$  has the same invariant distribution  $\pi$  as  $G'$  which means that the notion of analytic sparsity is unchanged. Further, if  $L$  and  $L'$  are the Laplacians of  $G$  and  $G'$ , respectively, then  $\langle f, Lf \rangle = \langle f, L'f \rangle$  for any  $f : V \rightarrow \mathbb{R}$ ; hence the notion of analytic boundary is also unchanged.

Given a reversible chain  $G$ , the operators  $K$ ,  $L$ , and  $H_t$  have a common orthogonal basis of eigenfunctions. We will write  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  for the eigenvalues of  $L$ ; note that the  $i$ th eigenvalue of  $K$  is  $1 - \lambda_i$  and the  $i$ th eigenvalue of  $H_t$  is  $\exp(-t\lambda_i)$ . All of our theorems which mention the eigenvalues  $\lambda_i$  hold also for non-reversible chains  $G'$ , with the  $\lambda_i$ 's being those for the associated reversible chain  $G$ .

Following [ABS10], our algorithm for the Analytic Small-Set Expansion problem (Theorem 1.4) breaks into two cases, depending on the “analytic nullity” of  $L$  (called “threshold rank” in [ABS10]):

**Definition 2.3.** We define  $\text{nullity}_\eta(L) = \#\{i : \lambda_i \leq \eta\}$ . Note that  $\text{nullity}_0(L)$  is the usual nullity.

**Remark 2.4.** Throughout we will present algorithms in the model of exact arithmetic. E.g., we will assume that given  $G$ , the eigenvalues and eigenfunctions of  $L$  can be computed exactly. We believe (but have not verified) that our results can be extended to standard computational models (e.g., Turing machines).

### 3 A new bound on the spectral profile

Here we give our new spectral criterion, based on the trace of the heat kernel, which ensures the existence of an analytically sparse function with small analytic boundary.

**Theorem 3.1.** Fix  $0 < \gamma \leq 1 \leq \Delta$  and suppose there exists  $t > 0$  such that

$$\text{tr}(H_t) - \frac{1}{\gamma} \text{tr}(LH_t) \geq \Delta. \quad (4)$$

Then in  $\text{poly}(n)$  time one can find  $g : V \rightarrow \mathbb{R}^{\geq 0}$  satisfying  $\mu[g] \leq 1/\Delta$  and  $\Phi[g] \leq \gamma$ .

*Proof.* Let  $\phi_x = \frac{1}{\pi(x)} \cdot 1_x$  for  $x \in V$ , so  $\mathbf{E}[\phi_x] = 1$ . Write  $\phi'_x = \sqrt{\pi(x)} \cdot \phi_x$ , so the collection  $(\phi'_x)_{x \in V}$  forms an orthonormal basis. Since trace is “the sum of the diagonal entries”, we have

$$\text{tr}(H_t) = \sum_{x \in V} \langle \phi'_x, H_t \phi'_x \rangle = \sum_{x \in V} \pi(x) \langle \phi_x, H_t \phi_x \rangle = \mathbf{E}_{x \sim \pi} \langle H_{t/2} \phi_x, H_{t/2} \phi_x \rangle.$$

Similarly,  $\text{tr}(LH_t) = \mathbf{E}_{x \sim \pi} [\langle H_{t/2} \phi_x, LH_{t/2} \phi_x \rangle]$ . Thus the assumption (4) implies

$$\mathbf{E}_{x \sim \pi} [\langle H_{t/2} \phi_x, H_{t/2} \phi_x \rangle - \frac{1}{\gamma} \langle H_{t/2} \phi_x, LH_{t/2} \phi_x \rangle] \geq \Delta.$$

Select (in  $\text{poly}(n)$  time) a particular  $x_0 \in V$  achieving at least  $\Delta$  in this expectation. We define  $g = H_{t/2}\phi_{x_0}$  and therefore we have

$$\langle g, g \rangle - \frac{1}{\gamma} \langle g, Lg \rangle \geq \Delta. \quad (5)$$

Note that  $g \geq 0$  since  $\phi_{x_0} \geq 0$  and  $H_{t/2}$  is positivity-preserving. Thus  $\|g\|_1 = \mathbf{E}[g] = \mathbf{E}[\phi_{x_0}] = 1$ . Further, from (5) we deduce  $\langle g, g \rangle \geq \Delta$ ; thus  $\mu[g] \leq 1/\Delta$  as desired. Finally, (5) certainly implies  $\langle g, g \rangle - \frac{1}{\gamma} \langle g, Lg \rangle \geq 0$ , which is equivalent to  $\Phi[g] \leq \gamma$ .  $\square$

A straightforward calculation now shows that if  $L$  has large analytic nullity then we can get good bounds from Theorem 3.1:

**Corollary 3.2.** *Fix  $0 < \gamma \leq 1$ . Let  $0 < \alpha \leq \frac{1}{3}$  and let  $k = \text{nullity}_{\alpha\gamma}(L)$ . Assume  $k \geq \frac{n^\alpha}{\ln n}$ . Then in  $\text{poly}(n)$  time one can find  $g : V \rightarrow \mathbb{R}^{\geq 0}$  satisfying  $\Phi[g] \leq \gamma$  and  $\mu[g] \leq 1/\Delta$ , where  $\Delta = \frac{k}{4n^\alpha}$ .*

*Proof.* We show that (4) from Theorem 3.1 holds with  $\gamma$ ,  $\Delta$ , and  $t = \frac{1}{\gamma} \ln n$ . We have

$$\text{tr}(H_t) - \frac{1}{\gamma} \text{tr}(LH_t) = \sum_{i=1}^n (1 - \frac{\lambda_i}{\gamma}) \exp(-t\lambda_i) = \sum_{i=1}^n (1 - \frac{\lambda_i}{\gamma}) n^{-\lambda_i/\gamma}. \quad (6)$$

The expression  $(1 - r)n^{-r}$  is decreasing for  $r \in [0, 1]$ ; for larger  $r$ , it attains its minimum at  $r = 1 + \frac{1}{\ln n}$ , where it has value  $-\frac{1}{e n \ln n}$ . Thus by distinguishing  $r = \frac{\lambda_i}{\gamma} \geq \alpha$  in (6) we may obtain

$$(6) \geq \#\{i : \lambda_i \leq \alpha\gamma\} \cdot (1 - \alpha)n^{-\alpha} - \#\{i : \lambda_i > \alpha\gamma\} \cdot \frac{1}{e n \ln n} \geq \frac{k}{n^\alpha} (1 - \alpha) - \frac{1}{e \ln n}.$$

Using  $\alpha \leq \frac{1}{3}$  and  $k \geq \frac{n^\alpha}{\ln n}$ , the above is indeed at least  $\Delta = \frac{k}{4n^\alpha}$ .  $\square$

Restating the parameters yields:

**Corollary 3.3.** *Let  $0 < \delta \leq 1$ . If there exists  $\alpha \leq \frac{1}{3}$  such that  $\text{nullity}_{\alpha\gamma}(L) \geq \frac{4}{\delta} n^\alpha$ , then in  $\text{poly}(n)$  time one can find  $g : V \rightarrow \mathbb{R}^{\geq 0}$  satisfying  $\mu[g] \leq \delta$  and  $\Phi[g] \leq \gamma$ .*

An alternative restatement of the parameters yields our main Theorem 1.3: simply take  $\alpha = \frac{1}{A \log_k n}$  and  $\gamma = A \lambda_k \log_k n$  in Corollary 3.2.

## 4 An algorithm for Analytic Small-Set Expansion

In [ABS10] it is shown that when  $L$  has small analytic nullity, one can find sparse sets by brute-force search through low-eigenvalue eigenspace. We present a very similar algorithm for finding analytically sparse sets.

**Lemma 4.1.** *Suppose there exists  $f : V \rightarrow \mathbb{R}$  with*

$$\mu[f] \leq \delta \leq 1/2, \quad \Phi[f] \leq \epsilon \leq 1/4.$$

*Let  $2\epsilon \leq \eta \leq 1$ . Then in time  $\exp(O(\text{nullity}_\eta(L) \log(\eta/\epsilon))) \cdot \text{poly}(n)$  one can find  $g : V \rightarrow \mathbb{R}$  satisfying*

$$\mu[g] \leq \delta + O(\epsilon/\eta + \sqrt{\delta\epsilon/\eta}) \leq O(\delta + \epsilon/\eta), \quad \Phi[g] \leq \eta.$$

**Remark 4.2.** It is also quite easy to show  $g$  will satisfy  $\Phi[g] \leq O(\sqrt{\epsilon/\eta})$ , which is useful if  $\eta \gg \epsilon^{1/3}$ . We will not need this parameter setting, so we omit the proof.

*Proof.* Let  $\psi_1, \dots, \psi_m$  be an orthonormal basis of eigenfunctions for  $L$ , corresponding to eigenvalues  $\lambda_1, \dots, \lambda_n$ . Without loss of generality, assume  $\|f\|_2 = 1$ . Write  $m = \text{nullity}_\eta(L)$  and write  $U$  for the dimension- $m$  subspace spanned by  $\psi_1, \dots, \psi_m$ . Express  $f = \sum_{i=1}^n c_i \psi_i$ , so  $\sum c_i^2 = 1$  by the orthonormality of the  $\psi_i$ 's. We have

$$\epsilon \geq \Phi[f] = \langle f, Lf \rangle = \sum_{i=1}^n \lambda_i c_i^2 \geq \sum_{i>m} \lambda_i c_i^2 \geq \eta \sum_{i>m} c_i^2.$$

In other words, if  $f_U$  denotes  $\sum_{i \leq m} c_i \psi_i$  then  $\|f - f_U\|_2^2 \leq \epsilon/\eta$  (which is at most  $1/2$  by the assumption on  $\eta$ ). If we define  $u \in U$  to be the unit vector  $f_U/\|f_U\|_2$ , it follows that

$$\|f - u\|_2 \leq \sqrt{2\epsilon/\eta}.$$

As in [ABS10] we can now consider all  $g$  in a  $.5\sqrt{\epsilon/\eta}$ -net for the unit sphere of  $U$ . The cardinality of this net is  $\exp(O(m \log(\eta/\epsilon)))$ . One such  $g$  will satisfy

$$\|u - g\|_2 \leq .5\sqrt{\epsilon/\eta} \quad \text{and hence} \quad \|f - g\|_2 \leq 2\sqrt{\epsilon/\eta}.$$

For this  $g$  we have

$$\|g\|_1 \leq \|f\|_1 + \|f - g\|_1 \leq \sqrt{\mu[f]} + \|f - g\|_2 \leq \sqrt{\delta} + 2\sqrt{\epsilon/\eta}$$

and hence  $\mu[g] \leq \delta + O(\epsilon/\eta + \sqrt{\delta\epsilon/\eta})$ , as desired. Since  $g$  is a unit vector in  $U$  we may also immediately conclude  $\Phi[g] \leq \eta$ .  $\square$

From Corollary 3.3 we know that if  $L$  has large analytic nullity then there is automatically an (easily findable)  $f : V \rightarrow \mathbb{R}$  which is analytically sparse and has small analytic boundary. On the other hand, if  $L$  has small analytic nullity, the above lemma can solve the Analytic Small-Set Expansion problem in not too much time. Combining these facts lets us prove our Theorem 1.4, restated here for convenience:

**Theorem 1.4.** *For any  $\alpha \leq \frac{1}{3}$  and  $C \geq 1$ , there exists an algorithm running in time  $\exp(O(n^\alpha \cdot \frac{1}{\delta} \log(C/\delta)))$  with the following guarantee: If there exists  $f : V \rightarrow \mathbb{R}$  with  $\mu[f] \leq \delta \leq 1/2$  and  $\Phi[f] \leq \epsilon \leq 1/4$ , the algorithm finds  $g : V \rightarrow \mathbb{R}$  with  $\mu[g] \leq \delta \cdot (1 + 1/C)$  and  $\Phi[g] \leq O(\frac{C^2}{\alpha\delta}) \cdot \epsilon$ .*

*Proof.* Set  $\gamma = \frac{B}{\alpha\delta} \cdot \epsilon$ ; we will eventually take  $B = O(C^2)$ . If  $\text{nullity}_{\alpha\gamma}(L) \geq \frac{4}{\delta}n^\alpha$  then from Corollary 3.3 we can find  $g$  with  $\mu[g] \leq \delta$ ,  $\Phi[g] \leq \gamma$  in  $\text{poly}(n)$  time; in fact, here we don't even need to assume the existence of  $f$ . Otherwise, Lemma 4.1 tells us that in time  $\exp(O(n^\alpha \cdot \frac{1}{\delta} \log(B/\delta)))$  we can find a  $g$  satisfying

$$\mu[g] \leq \delta + O(\frac{\delta}{B} + \sqrt{\frac{\delta^2}{B}}) = \delta \cdot (1 + O(1/\sqrt{B})), \quad \Phi[g] \leq \alpha\gamma \leq \gamma.$$

Thus the result follows by taking  $B = O(C^2)$ .  $\square$

## 5 The probability a random walk stays entirely within a set

In [OT12] the authors show that a  $t$ -step discrete time random walk starting from a random vertex in  $S \subseteq V$  stays entirely within  $S$  with probability at least  $\left(1 - \frac{\Phi[S]}{2}\right)^t$ . We give a proof of a similar result for continuous-time random walks using Markov chain methods.



**Theorem 1.7 restated.** For any  $\emptyset \neq S \subseteq V$  and real  $t > 0$ , let  $C(t, S)$  denote the probability that a continuous-time- $t$  random walk, started from a random  $\mathbf{x} \sim S$ , stays entirely within  $S$ . Then  $C(t, S) \geq \exp(-t\Phi[S])$ .

*Proof.* Let us define the operator  $K_S$  as follows: Given  $f : V \rightarrow \mathbb{R}$ ,  $K_S f(\mathbf{x}) = \mathbf{E}_{\mathbf{y} \sim \mathbf{x}}[f(\mathbf{y})1_S(\mathbf{y})]$ , where  $1_S$  is the indicator function for  $S$ . We then define  $L_S = I - K_S$  and  $H_{t,S} = \exp(-tL_S)$ . Also, let  $v_1, \dots, v_n$  be the eigenvectors of  $L_S$  (which are also eigenvectors of  $K_S$  and  $H_{t,S}$ ) and let  $\lambda_1, \dots, \lambda_n$  be the corresponding eigenvalues of  $L_S$ .

Define  $\phi'_S = \frac{1}{\sqrt{\pi(S)}} \cdot 1_S$ . We can then write  $\phi'_S = \sum_i c_i v_i$  for some constants  $c_i$ . Since  $\|\phi'_S\|_2 = 1$  it follows that  $\sum_i c_i^2 = 1$ .

First, we will show that  $\Phi[S] = \sum_i c_i^2 \lambda_i$ .

$$\begin{aligned} \Phi[S] &= \mathbf{Pr}_{\substack{\mathbf{x} \sim \pi \\ \mathbf{y} \sim \mathbf{x}}}[\mathbf{y} \notin S \mid \mathbf{x} \in S] \\ &= \mathbf{Pr}_{\substack{\mathbf{x} \sim \pi \\ \mathbf{y} \sim \mathbf{x}}}[\mathbf{y} \notin S \wedge \mathbf{x} \in S] / \mathbf{Pr}_{\mathbf{x} \sim \pi}[\mathbf{x} \in S] \\ &= \frac{1}{\pi(S)} \mathbf{E}_{\substack{\mathbf{x} \sim \pi \\ \mathbf{y} \sim \mathbf{x}}}[1_S(\mathbf{x})(1_S(\mathbf{x}) - 1_S(\mathbf{y}))] \\ &= \frac{1}{\pi(S)} \mathbf{E}_{\mathbf{x} \sim \pi}[1_S(\mathbf{x})(1_S(\mathbf{x}) - \mathbf{E}_{\mathbf{y} \sim \mathbf{x}}[1_S(\mathbf{y})])] \\ &= \frac{1}{\pi(S)} \mathbf{E}_{\mathbf{x} \sim \pi}[1_S(\mathbf{x})(I1_S(\mathbf{x}) - K_S 1_S(\mathbf{x}))] \\ &= \langle \phi'_S, L_S \phi'_S \rangle \\ &= \sum_i c_i^2 \lambda_i. \end{aligned}$$

Now we show that  $C(t, S) = \sum_i c_i^2 \exp(-t\lambda_i)$ . Let  $\mathbf{w}_0, \dots, \mathbf{w}_\tau$  be the states of a time- $t$  continuous-time random walk in  $G$ ; note that this is the same as a  $\tau$ -step discrete-time random walk, where  $\tau \sim \text{Poisson}(t)$ . Let  $\mathbf{W}$  denote the set of all states visited. Then:

$$\begin{aligned} C(t, S) &= \mathbf{Pr}[\mathbf{W} \subseteq S \mid \mathbf{w}_0 \in S] \\ &= \frac{1}{\pi(S)} \mathbf{E}[1_S(\mathbf{w}_0)1_S(\mathbf{w}_1) \dots 1_S(\mathbf{w}_\tau)] \\ &= \frac{1}{\pi(S)} \mathbf{E}_{\mathbf{x} \sim \pi} \mathbf{E}_{\tau \sim \text{Poisson}(t)}[1_S(\mathbf{x})K_S^\tau 1_S(\mathbf{x})] \\ &= \frac{1}{\pi(S)} \mathbf{E}_{\mathbf{x} \sim \pi}[1_S(\mathbf{x})H_{t,S}1_S(\mathbf{x})] \\ &= \langle \phi'_S, H_{t,S} \phi'_S \rangle \\ &= \sum_i c_i^2 \exp(-t\lambda_i). \end{aligned}$$

To complete the proof, we need to show that  $\sum_i c_i^2 \exp(-t\lambda_i) \geq \exp(-t \sum_i c_i^2 \lambda_i)$ . This follows immediately by the convexity of the exponential function and Jensen's inequality.  $\square$

## Acknowledgments

We thank James Lee, David Steurer, Yu Wu, and Yuan Zhou for helpful discussions.



## References

- [ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for Unique Games and related problems. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 563–572, 2010. 1, 1.1.1, 1.1.1, 2, 4, 4
- [Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. 1
- [AM85] Noga Alon and Vitali Milman.  $\lambda_1$ , isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985. 1
- [DSC96] Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Annals of Applied Probability*, 6(3):695–750, 1996. 2
- [GMT06] Sharad Goel, Ravi Montenegro, and Prasad Tetali. Mixing time bounds via the spectral profile. *Electronic Journal of Probability*, 11(1):1–26, 2006. 1.1.1, 1.2, 1
- [KL12] Tsz Chiu Kow and Lap Chi Lau. Finding small sparse cuts by random walk. In *Proceedings of the 16th Annual International Workshop on Randomized Techniques in Computation*, pages 615–626, 2012. 1, 1.1.2
- [LK99] László Lovász and Ravi Kannan. Faster mixing via average conductance. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 282–287, 1999. 1
- [LOT12] James Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order Cheeger inequalities. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, 2012. 1, 1, 1.1.1
- [LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, 2012. 1, 1.1.1
- [MT06] Ravid Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Foundations and Trends in Theoretical Computer Science*, 1(3):237–354, 2006. 2
- [OT12] Shayan Oveis Gharan and Luca Trevisan. Approximating the expansion profile and almost optimal local graph clustering. <http://arxiv.org/abs/1204.2021>, 2012. 1, 1.1.2, 5
- [RS10] Prasad Raghavendra and David Steurer. Graph expansion and the Unique Games Conjecture. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 755–764, 2010. 1.1.1
- [SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989. 1
- [Ste10] David Steurer. *On the Complexity of Unique Games and Graph Expansion*. PhD thesis, Princeton University, 2010. 1, 1.1.1